Amendments to the Specification:

Kindly delete the paragraph beginning on page 7, line 12 through page 8, line 9 and replace the same with the following amended paragraph.

In order to parameterize the feature log power spectrum, the frequency axis of the power spectrum is summarized by summing the energy within at least one predetermined frequency band. The summing may also be a weighted sum of energies, for example in dependence on frequency or in dependence on the energy itself. By way of example and not limitation, the predetermined frequency band can be at least one of the frequency bands 1-2 Hz, 3-15 Hz, and 20-150 Hz. The 1-2 Hz frequency band may be preferable to distinguish sounds with different rates of loudness changes as envelope modulations at very low frequencies are perceived as changes in loudness. Also musical tempo information is available from this frequency range. The 3-15 Hz frequency band may be preferable for classifying speech signals which contain prominent envelope modulations in the range of 3-15 Hz, which range corresponds to the syllabic rate. Other audio signals, such as music audio signals, have relatively fewer modulations in this range. The 20-150 Hz frequency band may be preferable to distinguish dissonant or rough sounds from consonant or smooth sounds as envelope modulations in the 20-150 Hz range are perceived as roughness, i.e. musical dissonance. Finally, the amount of energy within a predetermined frequency band may be divided by the average (DC) of subsequent values of the audio feature (i.e., subsequent values alone, not including preceding values) to yield a relative modulation depth. The average may be obtained by evaluating the 0 Hz energy in the feature power spectrum $|F|$. The result of this calculation is a further audio feature F_{mod} that can be used for classifying an audio signal. Another method to parameterize the feature log power spectrum is to transform the log-power spectrum $|F(f)|^2$ into at least one coefficient $C(m)$ using a discrete cosine transformation (DCT):

$$C(m) = \int_{f_a}^{f_b} \cos\left(\frac{(f - f_a)\pi m}{f_b - f_a}\right) \log \frac{|F(f)|^2}{|F(0)|^2} df \qquad (14)$$

in which f_a and f_b denotes the start and end frequency of a frequency band of interest. Usually, the upper frequency f_b is half the sampling frequency of f_s. Now, the coefficient *C(m)* is used as a further audio feature F_{mod}. *C(0)* denotes the total amount of modulations averages on a log scale, hence *C(0)* is related to the overall modulation depth. Due to the division of $|F(x)|^2$ by $|F(0)|^2$ the modulation depth parameter is independent of the signal level. Furthermore, it is noted that DCT coefficients may be highly uncorrelated which may be advantageous for audio classification. Also, it is noted that with an increasing number m of coefficients *C(m)*, more details of the feature log-power spectrum $|F(f)|^2$ are covered.